



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

E

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/822,976	03/30/2001	Joseph E. Pentheroudakis	M61.12-0317	8290
7590	10/22/2004		EXAMINER	
Joseph R. Kelly WESTMAN CHAMPLIN & KELLY International Centre, Suite 1600 900 South Second Avenue Minneapolis, MN 55402-3319			WOZNIAK, JAMES S	
			ART UNIT	PAPER NUMBER
			2655	
			DATE MAILED: 10/22/2004	

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No.	Applicant(s)	
	09/822,976	PENTHEROUDAKIS ET AL.	
	Examiner	Art Unit	
	James S. Wozniak	2655	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --
Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

- 1) Responsive to communication(s) filed on 30 March 2001.
- 2a) This action is FINAL. 2b) This action is non-final.
- 3) Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

- 4) Claim(s) 1-23 is/are pending in the application.
- 4a) Of the above claim(s) _____ is/are withdrawn from consideration.
- 5) Claim(s) _____ is/are allowed.
- 6) Claim(s) 1-23 is/are rejected.
- 7) Claim(s) _____ is/are objected to.
- 8) Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

- 9) The specification is objected to by the Examiner.
- 10) The drawing(s) filed on 3/30/2004 is/are: a) accepted or b) objected to by the Examiner.
 Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
 Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

- 12) Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) All b) Some * c) None of:
 1. Certified copies of the priority documents have been received.
 2. Certified copies of the priority documents have been received in Application No. _____.
 3. Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

- | | |
|---|---|
| 1) <input checked="" type="checkbox"/> Notice of References Cited (PTO-892) | 4) <input type="checkbox"/> Interview Summary (PTO-413) |
| 2) <input type="checkbox"/> Notice of Draftsperson's Patent Drawing Review (PTO-948) | Paper No(s)/Mail Date. _____ |
| 3) <input checked="" type="checkbox"/> Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08)
Paper No(s)/Mail Date <u>5/8/2003</u> . | 5) <input type="checkbox"/> Notice of Informal Patent Application (PTO-152) |
| | 6) <input type="checkbox"/> Other: _____ |

DETAILED ACTION

Claim Objections

1. **Claim 23** objected to because of the following informalities: “claim 23” in line 1, should be corrected to “claim 22.”

Appropriate correction is required.

Claim Rejections - 35 USC § 102

2. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

A person shall be entitled to a patent unless –

(e) the invention was described in (1) an application for patent, published under section 122(b), by another filed in the United States before the invention by the applicant for patent or (2) a patent granted on an application for patent by another filed in the United States before the invention by the applicant for patent, except that an international application filed under the treaty defined in section 351(a) shall have the effects for purposes of this subsection of an application filed in the United States only if the international application designated the United States and was published under Article 21(2) of such treaty in the English language.

3. **Claims 1-3, 5-7, 10, and 11** are rejected under 35 U.S.C. 102(e) as being anticipated by Bond et al (*U.S. Patent: 6,539, 348*).

With respect to **Claim 1**, Bond discloses:

Receiving the input string (*Col. 3, Lines 9-10*);

Segmenting the input string into one or more proposed tokens (*Col. 3, Lines 21-29*);

Validating the proposed tokens by submitting the proposed tokens to a linguistic knowledge component to determine whether the proposed tokens represent linguistically meaningful units (*Col. 3, Lines 29-48*); and

If not, re-segmenting the input string into one or more different proposed tokens (*Col. 3, Lines 45-61*).

With respect to **Claim 2**, Bond recites:

Accessing segmentation criteria arranged in a predetermined hierarchy of segmentation criteria, and segmenting based on the segmentation criteria in an order based on the hierarchy (*Col. 10, Lines 11-24*).

With respect to **Claim 3**, Bond discloses:

Accessing language-specific data containing a portion of the segmentation criteria (*Col. 7, Lines 19-23*).

With respect to **Claim 5**, Bond discloses:

Validating and re-segmenting until all characters in the input string have been validated or until the predetermined hierarchy of segmentation criteria has been exhausted (*Col. 2, Lines 3-11*).

With respect to **Claim 6**, Bond recites:

Accessing the lexicon to determine whether it contains the proposed tokens (*Col. 3, Lines 29-48*).

With respect to **Claim 7**, Bond discloses:

Invoking the morphological analyzer to convert a form of the proposed tokens to a morphologically different form (*Col. 3, Lines 45-50*); and

Accessing the lexicon to determine whether it contains the morphologically different form of the token (*Col. 3, Lines 50-52*).

Claim 10 contains subject matter similar to Claims 1 and 7, and thus, is rejected for the same reasons.

With respect to **Claim 11**, Bond recites:

Repeating the steps of proposing a subsequent segmentation and submitting the subsequent segmentation to the linguistic knowledge component until the portion of the input string is validated or the portion of the input string has been segmented according to a predetermined number of segmentation criteria (*Col. 10, Lines 11-24*).

Claim Rejections - 35 USC § 103

4. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

5. **Claims 4, 8, 9, 12, and 17-19** are rejected under 35 U.S.C. 103(a) as being unpatentable over Bond et al in view of Carus (*U.S. Patent: 5,890,103*).

With respect to **Claim 4**, Bond teaches the natural language processing system utilizing a dictionary lookup-process in determining a proper sentence segmentation format, as applied to Claim 1. Bond does not specifically suggest language-dependent punctuation data in a segmentation process, however, Carus recites:

Accessing a precedence hierarchy of punctuation in the language-specific data, the precedence hierarchy being arranged based on binding properties of the punctuation in the precedence hierarchy, and segmenting the input string based on the punctuation in an order based on the precedence hierarchy (*segmentation based on punctuation placement for specific languages, Col. 42, Lines 10-37*).

Bond and Carus are analogous art because they are from a similar field of endeavor in natural language processing. Thus, it would have been obvious to a person of ordinary skill in the art, at the time of invention, to modify the teachings of Bond with the use of language specific segmentation rules based on punctuation placement as taught by Carus to implement higher level linguistic processing (*Carus, Col. 2, Lines 9-16*) in order to prevent incorrect segmentation by identifying special characters (*Carus, Col. 39, Lines 23-28*) that could have different meanings for a specific language based upon character location.

With respect to **Claim 8**, Bond teaches the natural language processing system utilizing a dictionary lookup-process in determining a proper sentence segmentation format, as applied to Claim 1, while Carus teaches the use of language specific segmentation rules based on punctuation placement as applied to Claim 4.

Bond and Carus are obvious in combination for the reasons noted with respect to Claim 4.

Claim 9 contains subject matter similar to Claim 5, and thus, is rejected for the same reasons.

With respect to **Claim 12**, Bond teaches the natural language processing system utilizing a dictionary lookup-process in determining a proper sentence segmentation format, as applied to

Claim 10. Bond does not specifically suggest segmenting an input string at white spaces, however such a segmenting technique is well known and commonly used in the art as a more basic means of parsing input text as is evidenced by Carus (*Col. 13, Lines 62-64*).

Bond and Carus are analogous art because they are from a similar field of endeavor in natural language processing. Thus, it would have been obvious to a person of ordinary skill in the art, at the time of invention, to modify the teachings of Bond with the means of segmenting an input text string at white spaces as taught by Carus to provide basic initial segmentation of an input text, based on white spaces, for further linguistic analysis.

With respect to **Claim 17**, Carus discloses the means of segmenting an input text string at white spaces, as applied to Claim 12.

With respect to **Claim 18**, Bond additionally discloses:

Determining whether the token contains either all alpha characters or all numeric characters; and if so, indicating that the token cannot be further segmented and will be treated as an unrecognized word (*acronym containing all capital letters marked as a unknown word and assigned a token, Col. 3, Lines 45-61*).

With respect to **Claim 19**, Carus further recites:

Determining whether the token includes final punctuation; and if so, segmenting the token into a subtoken by splitting off the final punctuation (*Jones', Col. 41, Lines 1-10*).

6. **Claims 13 and 14** are rejected under 35 U.S.C. 103(a) as being unpatentable over Bond et al in view of Carus, and further in view of Grefenstette (*U.S. Patent: 6,289,304*).

With respect to **Claim 13**, Bond in view of Carus teaches the natural language processing system utilizing a dictionary lookup-process in determining a proper sentence segmentation format and a means of segmenting an input text string at white spaces, as applied to Claim 12. Bond in view of Carus does not teach the detection and segmentation of emoticons, however, Grefenstette discloses:

Determining whether invalid tokens contain any of a predetermined plurality of multi-character punctuation strings or emoticons and, if so, segmenting the tokens into subtokens based on the multi-character punctuation strings or emoticons (“smileys,” *Col. 4, Line 62- Col. 5, Line 4*).

Bond, Carus, and Grefenstette are analogous art because they are from a similar field of endeavor in natural language processing. Thus, it would have been obvious to a person of ordinary skill in the art, at the time of invention, to modify the teachings of Bond in view of Carus with the means of parsing emoticons in a text string as taught by Grefenstette in order to increase natural language processing system capabilities by implementing a means for recognizing and segmenting emoticons which would otherwise have no meaning in a traditional lexicon.

With respect to **Claim 14**, Carus additionally recites:

Determining whether invalid tokens contain punctuation marks; and if so, segmenting the tokens into subtokens according to a predetermined precedence hierarchy of punctuation (*detecting an apostrophe within a text string and segmenting text based on apostrophe location, Col. 40, Lines 1-50, and Col. 42, Lines 10-37*).

7. **Claims 15, 16, and 21-23** are rejected under 35 U.S.C. 103(a) as being unpatentable over Bond et al in view of Carus, further in view of Grefenstette, and in yet further view of Malsheen et al (*U.S. Patent: 5,634,084*).

With respect to **Claim 15**, Bond in view of Carus, and in further view of Grefenstette teaches the natural language processing system capable of detecting punctuation marks within a token, as applied to Claim 14. Bond in view of Carus, and in further view of Grefenstette does not teach determining whether a token contains both alpha and numeric characters and segmenting a string containing such characters at alpha-numeric boundaries, however Malsheen suggests:

Determining whether invalid tokens contain both alpha and numeric characters; and if so, segmenting the tokens into subtokens at boundaries between the alpha and numeric characters in the tokens (*parsing syllables that can consist of a combination of letters and numbers, Abstract*).

Bond, Carus, Grefenstette, and Malsheen are analogous art because they are from a similar field of endeavor in linguistic processing. Thus, it would have been obvious to a person of ordinary skill in the art, at the time of invention, to modify the teachings of Bond in view of Carus, and in further view of Grefenstette with the ability to determine whether a token contains both alpha and numeric characters and segment a string containing such characters as taught by Malsheen in order to improve natural language processing system capabilities by parsing alpha-numeric words which would not be separated using conventional text processing (*Malsheen, Col. 2, Lines 53-59*).

With respect to **Claim 16**, Bond further recites:

Reassembling previously segmented subtokens (*producing a revised token, Col. 3, Lines 45-61*).

Claim 21 contains subject matter similar to Claim 13, and thus, is rejected for the same reasons.

With respect to **Claim 22**, Carus additionally recites:

Determining whether the token includes one or more edge punctuation marks; and if so, segmenting the token into subtokens by splitting off the one or more edge punctuation marks according to a predetermined edge punctuation precedence hierarchy ('twas, *Col. 40, Lines 1-23*).

With respect to **Claim 23**, Carus additionally recites:

Determining whether the token includes one or more internal punctuation marks, internal to the tokens; and if so, segmenting the token into subtokens based on the one or more internal punctuation marks according to a predetermined internal punctuation precedence hierarchy (*male/female token, Col. 37, Line 63- Col. 38, Line 6*).

8. **Claim 20** is rejected under 35 U.S.C. 103(a) as being unpatentable over Bond et al in view of Carus, and further in view of Malsheen et al.

With respect to **Claim 20**, Bond in view of Carus teaches the natural language processing system capable of detecting final punctuation marks within a token, as applied to Claim 19, while Malsheen teaches the ability to determine whether a token contains both alpha and numeric characters and segment a string containing such characters as applied to Claim 15.

Bond, Carus, and Malsheen are obvious in combination for the reasons noted with respect to Claim 15.

Conclusion

9. The prior art made of record and not relied upon is considered pertinent to applicant's disclosure:

- Parra (*U.S. Patent: 5,870,700*)- teaches a means of additional language specific natural language token processing.
- Williams (*U.S. Patent: 5,963,742*)- teaches the use of multiple subparsers that each differently interpret tokens.
- Newsted (*U.S. Patent: 6,016,467*)- teaches the use of punctuation characters to separate tokens.

10. Any inquiry concerning this communication or earlier communications from the examiner should be directed to James S. Wozniak whose telephone number is (703) 305-8669 and email is James.Wozniak@uspto.gov. The examiner can normally be reached on Mondays-Fridays, 8:30-4:30.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Doris To can be reached at (703) 305-4827. The fax/phone number for the Technology Center 2600 where this application is assigned is (703) 872-9306.

Any inquiry of a general nature or relating to the status of this application or proceeding should be directed to the technology center receptionist whose telephone number is (703) 306-0377.

James S. Wozniak
10/7/2004



SUSAN MCFADDEN
PRIMARY EXAMINER